TOWARDS THE USE OF A NOVEL METHOD: The First Experiences on Measuring the Cognitive Load of Learned Programming Skills

Assist. Prof. Dr. Murat Pasa UYSAL,
Turkish Military Academy
Ankara, TURKEY

ABSTRACT

Teaching object-oriented programming (OOP) is a difficult task, especially to the beginners. First-time learners also find it difficult to understand. Although there is a considerable amount of study on the cognitive dimension, a few study points out its physiological meaning. Moreover, it has been suggested that neuroscientific studies and methods can enable educational researchers gain an insight into brain and cognitive processes as well. Therefore, this experimental study explored the previously learned OOP skills in terms of cognitive load. By using the functional near-infrared spectroscopy (fNIRS) method, we measured the cognitive load of participants when executing OOP tasks. The average oxygenation changes in prefrontal cortex of the brain represented their total cognitive response to a set of OOP tasks. There were two research questions investigated by this study. The first research question explored whether the average oxygenation changed according to the participants' academic achievements and task completion status. The second research question was for identifying the instant changes in the oxygenations to find out which programming tasks were more contributing to the oxygenation. Later, we compared the findings with experts' judgments. We observed that the fNIRS system was an effective and promising technology for monitoring cognitive tasks both in classrooms and in experimental environments.

Keywords: Cognitive load, object-oriented programming, fNIR spectroscopy.

"I, of course, am not a neuroscientist, and even I have never dreamed of engaging with this type of discipline. However, my exciting exploratory journey has started after participating to a seminar (Modsimmer, 2010) on the functional near-infrared spectroscopy (fNIRS) method. M.P. Uysal

INTRODUCTION

We are living in an information age, and most of the complex problems, which our society faces cross the disciplinary boundaries. It is often required to integrate different knowledge domains, and collaborate effectively to solve the problems. The possible solutions can be produced through an interdisciplinary approach, which provides comprehensive understanding of issues and challenges. Collaboration of neuroscience and education could be given as an example, and a considerable amount of publications is advocating the integration of neuroscientific and educational research (Griffin and Case, 1997; Berninger and Corina, 1998; OECD, 2002; Campbell, 2006; Nes and De Lange, 2007; Szucs, Soltész, Jármi, and Csépe, 2007; Nes, 2011). We believe that neuroscientific research and methods can enable educational researchers gain an insight into brain and cognitive processes. On the other hand, behavioral subjects that educational researchers study can help understand the nature of the changes in human brain, which often result as a function of learning process (Nes, 2011). Thus, the Cognitive Load Theory (CLT) has been an intermediate research between these disciplines.

The term cognitive load, which is one of the indicators for cognitive performance, represents the load that imposed on the cognitive system when performing a task (Gog, Kester, Nievelstein, Giesbers, & Paas, 2009). CLT identifies three types of cognitive load: intrinsic cognitive load, extraneous cognitive load and germane cognitive load (Sweller, Merriënboer, Jeroen, & Paas, 1998). Its primary assumption is that learners have a limited cognitive processing capacity for acquiring any knowledge and skill. Therefore, instructional designers should use appropriate strategies, and control learning environment for an effective knowledge transfer. An instruction should be designed so that learners use this limited capacity while making the information process easier in working memory (WM). With in the framework of CLT, the "intrinsic cognitive" load indicates the inherit complexity of instructional materials and it is not generally manipulated by instructional designs (Sweller et al., 1998). The extraneous cognitive load is generated by an ineffective instruction resulting from poorly designed instructional activities. The "germane cognitive load" represents the effort that contributes to the construction of new schemas and knowledge acquisition (Gog et al., 2009).

The general consent is that teachers and instructional designers should use the methods for increasing the germane cognitive load while trying to reduce the intrinsic and extraneous cognitive load. Therefore, the studies done on the cognitive load generally cover different research topics depending on these prescriptions. Consequently, the cognitive load measurement has been one of the issues for instructional designs guided by CLT.

PROBLEM

OOP learners usually find programming concepts and techniques difficult to envision. When teaching OOP to novices with minimally guided instruction, the working memory resources of learners can quickly become overwhelmed by task complexity (Mayer and Moreno, 2003; Kirschner, Sweller and Clark, 2006; Artino, 2008). Furthermore, OOP is a matter of thinking in object-oriented way as well as understanding underlying conceptual framework. While this paradigm is close to Human's natural perception of the real world, it has often been difficult to introduce novice learners to its perspective.

The studies done on teaching OOP can be grouped into three categories: (1st) the educational tools for OOP learners, (2nd) the instructional approaches to teaching OOP, (3rd) and learners' characteristics and attitudes to OOP (Xinogalos, 2010). Although, the instructional tools and methods of OOP posses the primary and effecting parameters of teaching OOP, a special notice should be given to the cognitive dimension. Learning OOP skills require fundamental cognitive processes, and the cognitive capacity plays a mediating factor through out a learning process. Therefore, many studies underlined the importance of WM in computer problem solving (Merriënboer and Paas, 1990; Sohn and Doane, 1997; Doane, Sohn, McNamara and Adams, 2000; Soan and Doane 2003).

In general, cognitive load related studies condense on the extraneous and/or germane cognitive load rather than the intrinsic cognitive load. The primary reason would be their manipulability by the instructional settings, tools, and strategies. However, three types of cognitive load are additive in a learning process, and they have an integral effect on cognitive performances. An effective knowledge transfer can occur if the total cognitive load were managed consciously during the instructional processes. Thus, learning is improved if an instruction is designed regarding the interaction of these different cognitive loads. Therefore, we should seek for the ways to control or monitor the cognitive load while applying the principles of CLT to the instructional settings. As the prior knowledge plays an important role, the cognitive dimension is also an affecting factor when learning OOP skills (White and Sivitanides, 2005). The intrinsic cognitive load of the core concepts of OOP is high especially for novice learners.

Therefore, understanding how to measure or observe the effects of cognitive load is a fundamental challenge.

With this study, we primarily intended to explore how effective previously learned OOP skills and knowledge were processed in WM. This paper presents the results of the experimental study that adopted the functional near-infrared spectroscopy (fNIRS) method as a relatively new technique. We measured the participants' average oxygenation changes in the prefrontal cortex of brain when executing experimental OOP tasks. These values represented their total cognitive response to a set of OOP tasks. Our study tried to answer two research questions. The first one was to find out whether average oxygenation would change according to the task completion status and academic achievement test. The second was to compare the expert level judgments of OOP tasks with the average oxygenations.

We could use traditional methods, self-reported tools, or other similar techniques to measure the cognitive load of OOP learners. However, these techniques could not serve the purpose of our research. In essence, learning means a permanent physiological change in Human brain, and the current literature of learning OOP lacks this dimension. This is one of the reasons for why we explored the cognitive performance of OOP learners using the fNIRS method. We present our initial experiences in the next sections.

COGNITIVE LOAD MEASUREMENT AND THE fNIRS SYSTEM

Sweller et al. (1998) specifies three mental effort measurement techniques. These are the task & performance-based measurement, the subjective measurement, and the physiological measurement. Gog et al. 2009 considers the subjective rating scales as an appropriate approach for classroom environments. On the other hand, Ayaz, Willems, Bunce, Shewokis, Izzetoglu, Hah, Deshmukh and Onaral (2010) classify the direct and indirect cognitive load assessment tools into four main groups.

The subjective assessment methods that use self-reported rating scores take place in the first group. Behavior and performance measures of the participant are included in the second. The methods based on physiological measures, such as heart rate, respiration rate and eye movements are in the third group.

Finally, the tools in the fourth group of monitor the cognitive activities and they use the signals directly from the brain. The human brain goes through a number of physiological changes while responding to environmental stimuli (Carrióna, Izzetoglu M., Izzetoglu K., Rodrígueza, López, Barroso, & Morales, 2010).

Some of these physiological events are related with brain activities. Different optical techniques have been used for the assessment of these events and activities. Magnetoencephalography (MEG), magnetic resonance imaging (MRI) and positron emission tomography (PET) are prominent techniques for studying the brain activities. However, these techniques possess important constraints. For example, the individuals may subject to potentially harmful equipments in some circumstances. These equipments and materials are highly expensive. Finally, these methods confine subjects to restricted positions since their equipments are sensitive to motion artifacts (Ayaz et al., 2010).

It was Jobsis (1977) who initially reported that the light in the near-infrared range (NIR), diffusing through scalp and skull, could be used to measure brain activities. The changes in electrochemical activity and in blood levels influence the functional state of tissue, and affect the optical properties of the human brain. The injected photons follow different paths when NIR range of the spectrum is introduced at the scalp.

The layers of head, which are skin, skull and brain, absorb some of these photons, and the others follow different pattern as a result of the scattering effect of tissue (Villringer and Chance 1997). If a special optical apparatus detects the scattered back photons, it can be seen that oxygenated (HbO2) and deoxygenated hemoglobin (Hb) are the primary absorbers in the NIR range (Bozkurt, Rosen, Rosen and Onaral, 2005). When the spectrum of light is analyzed, backscattered photons are interpreted as the changes in blood chromophores since lipid and water are relatively transparent to NIR light. Therefore, the information about blood volume and tissue oxygenation may be an evidence of hemodynamic activity in the dorsolateral prefrontal cortex. This part of the brain is the focus of our experimental study.

The fNIR spectroscopy system is comprised of a control box, a light source and detector, and a pre-configured system including the processing software (Figure 1). The headband attached to a subject's forehead receives the light reflected from the tissue, and then it produces 16 measurement locations (voxels) per wavelength through the 4 light sources and 10 detectors. The system simultaneously gauges the forehead oxygenation changes representing the hemodynamic activities. Using a continuous wave spectroscopy applied to tissue at constant amplitude; it calculates the values of oxygenated and deoxygenated hemoglobin molecules relative to a baseline. The special software acquires, visualizes, and records the raw data through out the experiment (Appendix-2). The fNIRS system is accepted as a portable and functional research tool for monitoring cognitive tasks in real life situations as well as in experimental environments (Izzetoglu et al., 2004).

Figure: 1
The fNIR system and the experimental environment

It is important to note that functional hemodynamic changes (the increase or decrease of neural activities in prefrontal cortex) can be interpreted as the result of learning (Leon, Izzetoglu M., Izzetoglu K., Rodrigez, Damas, Martin and Morales, 2010). They are also associated with the WM and executive processes.

Selected Studies on fNIRS Method

Lenf, Espina, Elwell, Athanasiou, Delpy, Darzi and Yang (2011) made a systematic literature review of fNIR technique, and they suggest that the studies can be broadly divided into two categories: (1st) technical advances and aspects of the fNIRS method; (2nd) its application to different research areas as an effective and noninvasive tool. We included only the cognitive load related studies to give an idea about fNIRS method. From a technical aspect, Izzetoglu M., Izzetoglu K., Bunce, Ayaz, Devaraj and Onaral B. (2005) present a broad description of fNIR system, such as its instrumentation, signal processing, data analysis and its visualization. They underline that it can be integrated with other physiological and neurobehavioral measures, eye tracking, pupil reflex, heart rate variability, respiration etc. Izzetoglu M., Bunce, Izzetoglu K., Onaral and Pourrezai (2007) describe the working principles of fNIR and the use of modified Beer-Lambert Law for signal extraction from raw fNIR measurements. They summarize the merits of optical imaging in augmented cognition.

Concerning the physiological measurement of cognitive load, Izzetoglu, Bunce, Onaral and Chance (2004) carried out an experiment on augmented cognition. Their experimental protocol uses a complex command and control task resembling a video game. Their primary hypothesis is that blood oxygenation in the prefrontal cortex would rise with the increasing task load and it would demonstrate positive correlation with performance measures.

Their results suggest a reliable, positive association between the cognitive load and the increase in oxygenation under circumscribed conditions. Similar to this study, Ayaz, Willems, Bunce, Shewokis, Izzetoglu, Hah, Deshmukh and Onaral (2010) use fNIR system to obtain the measures of cognitive load. In their study, air traffic controllers manage the scenarios under typical and emergent conditions as they complete n-back task with an increasing difficulty.

The findings are parallel to the earlier studies in which the blood oxygenation changed. Leon et al. (2010) focuses on the physiological effects of repetition on learning and WM, and they use adaptation of Luria's Memory Word-Task. They observe attenuation of stimulus-evoked neural activity in prefrontal neurons, and efficient verbal learning is mediated by the neural repetition suppression mechanism.

METHOD

As we stated before, the purpose of the study was to explore how effective previosly learned OOP skills and knowledge were processed in WM. This cognitive performance was associated with the participants' cognitive load measured as the average oxygenation changes. These values represented their total cognitive response to a set of OOP tasks. Our study had two phases (Table: 1). The first phase covered a 13-week graduate course aiming to teach the fundamentals of OOP.

Although this phase is not in the scope of our study, it ensured that the participants acquired the required skills and knowledge of OOP, and they were ready for the fNIRS phase of study. Afterwards, they took a supplementary 3-hour lecture covering the topics directly related to the fNIRS experiment.

Table: 1
The Research Design

Phase-1 (Prior to study)		*Phase-2 (Experimental study)*			
OOP Course	Programming Tool	Supplementary Lecture	Programming Tool	Post-test	Measurement
13 weeks	BlueJ	3 hours	BlueJ	Academic achievement	Cognitive load (fNIRS)

The same instructor gave both the OOP course and the lecture. An academic achievement test determined whether these skills and knowledge retained by the learners following the lecture. Prior to the fNIRS study, the experimenter presented some previous applications of fNIRS as a noninvasive tool, and eleven male volunteers eventually participated to the fNIRS experiement.

The participants signed consent forms, and they were informed that leaving the fNIRS experiment was allowed at anytime.The participants' average oxygenation change represented their cognitive response to a set of OOP tasks. Aligned with the study goals; we formulated the research questions as follows:

R.Q-1. In terms of cognitive load represented by the average oxygenation measures,

- **Is there any significant difference between the participants according to their achievement tests?**
- **Is there any significant difference between the participants according to their task completion status?**

R.Q-2. What are the subject matter experts' evaluations on the difficulty levels of the experimental OOP tasks? Is there any relationship between the average oxygenation measurements and the expert evaluations?

Procedure

The participants took a 3-hour supplementary lecture on the basic concepts of OOP, which was subsequent to a 13-week OOP course. The primary purpose was to wrap up the topics related to the experimental OOP tasks. The core concepts, such as inheritance, polymorphism, encapsulation, and the worked examples were given during this lecture. The open-ended academic achievement test determined whether the specific skills and competencies were at a desired level.

The test had two sections. The participants were required to give correct descriptions of OOP concepts at the first section.

As to the second, they wrote a complete Java program applying these concepts. They were graded based on the knowledge on the OOP concepts, and the Java applications developed during the test. Afterwards, this OOP academic achievement test was transformed to an fNIRS experiment protocol.

It was to make sure that the observed differences in the cognitive load resulted from the previously learned OOP tasks, not from unknown or newly encountered tasks. We also converted this experimental protocol to an evaluation scale. The experienced OOP instructors evaluated difficulty levels of these experimental tasks (Appendix 1).

The duration for each participant to complete the fNIRS session was between 40 and 60 minutes including the time needed for adjustments and experimental procedures. However, they were supposed to finish the experimental OOP tasks in 20 minutes. Initially, they were seated in front of their notebooks, and the experimenter read the OOP tasks to the participants, and answered their questions before the experiment. This was to make sure that everything was clarified before fNIRS measurements. A special notice was given such that the participants were relaxed and felt comfortable during the measurements to avoid possible motion artifacts. Therefore, the task document was placed at an eyesight level to enable the subjects to read in a suitable position. The measurements began with a baseline while participants were resting and relaxing (Figure: 1). The required data was collected from the sixteen-channel fNIR system while subjects were completing their tasks. At the same time, the experimenter observed each session and used manual markers to record the exact time of actions or possible changes when executing the OOP tasks. These recordings were later compared with the experts' judgments on the corresponding experimental tasks. Their sessions ended when the allowed time expired, and only six participants were able to complete all of the tasks.

FINDINGS AND DISCUSSION

The special software (Appendix 2) recorded two wavelengths, and each of the 16 voxels made 48 measurements for each sampling period.

Table: 2
The descriptive data of the study

Participants	Academic achievement	Task completion	Average oxygenation
P-1	81	Yes	0,0003
P-2	93	No	1,4956
P-3	92	No	2,3001
P-4	82	No	0,8804
P-5	92	Yes	-0,3160
P-6	91	Yes	-3,5953
P-7	87	Yes	-0,0277
P-8	80	No	1,4011
P-9	94	Yes	-1,2581
P-10	92	Yes	-0,7209
P-11	88	No	1,2573

The averaged oxygenation and rate of oxygenation were calculated by a testing and analysis platform (Izzetoglu et al., 2004).

The average oxygenation value for each subject was extracted from a set of approximately (48 measurements x 16 voxels x 2400) lines of raw data after series of procedures. Then, the collected data was low-pass filtered to eliminate the artifacts such as the equipment noise, heart pulsation, respiration and motion artifacts.The average oxygenation change of each participant took place in a scale between positive or negative numbers depending on individual data range. The neural activities in prefrontal cortex part of human brain are defined as the hemodynamic changes during cognitive processes, and they are associated with the WM and executive processes (Leon, Izzetoglu M., Izzetoglu K., Rodrigez, Damas, Martin and Morales, 2010).

The literature review also shows that the greater expertise is associated with lower oxygenation (less neural activity) at different levels or types of a cognitive task (Izzetoglu et al., 2004; Bunce et al., 2006; Izzetoglu et al., 2007; León-Carrióna et al., 2010; Ayaz et al., 2012). Therefore, it is possible to interpret the oxygenation values as:

- (1st) the complete cognitive response of learners to the experimental OOP tasks,
- (2nd) the total intrinsic load resulting from nature of the OOP tasks.

We used SPSS v.16 software for the statistical analysis processes to answer the research questions. Table 2 presents the descriptive data of the study.The purpose of this research question was to determine whether there were relationships, amongst the academic test scores, the task completion status and average oxygenation changes. The supplementary lecture and the academic achievement tests indicated that the learners acquired the OOP knowledge and skills (Table: 2). However, some of the participants (S-1, S-5, S-6, S-7, S-9 and S-10) completed the set of OOP tasks, while the others (S-2, S-3, S-4, S-8 and S-11) were not able to accomplish. The statistical methods, such as the correlation (Spearman's rho and Kendall’s tau), regression and Mann Whitney tests were used to explore possible relationships. Table: 3 ummarizes the results.

Table: 3
The statistical analyses of average oxygenation change (aoc), task completion status (tcs) and academic achievement test (aat)

Variables	Statistical Method	Coefficients / z	Significance
aoc / tcs	Correlation	rho = -,866 tau = -,739	,001 ,006
aoc / aat	Correlation	rho = -,197 tau = -,187	,562 ,432
aoc / tcs	Regression	r2 = ,613	,004
aoc / tcs	Mann-Whitney	z = -2,739	,004

R-1.a. In terms of cognitive load represented by the average oxygenation measures, is there any significant difference between the participants according to their achievement tests?

The OOP section of the academic achievement test and the fNIRS experiment protocol were similar. This was to explore previously learned OOP tasks, and to observe the differences in oxygenation levels, which were possibly resulted from processing the information in WM.

The achievement tests given subsequent to the supplementary lecture showed that the learning outcomes were achieved, and the learners mastered OOP skills and knowledge directly related to the fNIRS experiment (Table 2). However, the results of statistical tests do not show a significant correlation between the average oxygenation changes and academic achievement tests (aoc/aat) (rho= -,197; p = ,562). In other words, the participants' cognitive load, which was represented by the average oxygenation measures, did not change significantly according to the academic achievements. It seemed that the learners acquired the OOP skills and knowledge necessary for the fNIRS experiments, but they could not reflect their competencies to their cognitive performances. The participants with high-test results did not have low cognitive loads in terms of average oxygenations or vice versa. As a result, the greater expertise could not be associated with lower oxygenation contrary to our expectations. One of the primary reasons for this result would depend on how the OOP facts and rules were processed during both the OOP course and lecture. It is generally assumed that learning outcomes are achieved if learners can recall and use desired information according to the accepted standards.

The duration for a task is usually adjusted to a limited time to discriminate learners' performances. This was also the same for our experiment. In addition, the experimental tasks were exactly same with the tasks in the achievement tests. The literature review suggests that information is stored in stable forms and structures in the long-term memory (LTM), and reliable access or retrieval is possible by means of information cues. To perform a complex cognitive task as in OOP, learners have to maintain access to large amounts of information in the LTM. How this information is coded and processed, and how the retrieval cues are formed during instruction directly affects the quality and effectiveness of cognitive performance.

In this study, the scarcity in retrieval cues, which had to be associated with the OOP items in LTM, possibly caused a bottleneck for information retrieval. In the traditional model of human memory, immediate recall yields items that are directly retrieved from a temporary short-term memory, or the items retrieved by the cues from a more durable storage in LTM (Atkinson and Shiffrin, 1968; Waugh and Norman, 1968).

Many theorists (Shiffrin, 1976; Schneider and Detweiler, 1987) proposed that recalling should be mediated by the cues in a learning context, and schemas must be formed by using proper strategies (Ericsson and Kinstch, 1995). Merriënboer and Sweller (2005) also indicate, "Human expertise comes from knowledge stored in the schemata, not from an ability to engage in reasoning with many elements that have not been organized in LTM". However, novice OO programmers generally lack effectively organized schemas necessary to process programming tasks. Without instructional guidance, they can quickly have variant cognitive load, if not an inefficient learning, an unstructured information store will occur in LTM. Therefore, instructional settings should assist learners in forming the OPP context with cues and memory structures (Sweller et al., 1998). Before the fNIRS measurements, both the OOP course and supplementary lecture adopted a classical instructional design as in many programming courses. We believe that the general notion of the course was away from providing OOP learners with necessary retrieval cues. Since the interrelated information structures of OOP were not presented consciously, the learners processed the information according to their learning styles and cognitive preferences.

Consequently, this might have affected the learners' cognitive performance, and their cognitive load possibly uncorrelated with their academic achievement test results.

R-1.b. In terms of cognitive load represented by the average oxygenation measures, is there any significant difference between the participants according to their task completion status?

The previous research question could not identify any relationship between the academic achievement test and the cognitive load. However, the statistical tests show a significant and negative correlation between average oxygenation change and task completion status (aoc/tcs) (rho =-,866; p = ,001). The regression and Mann-Whitney test results also show that the participants' OOP task completion status can be predicted according to their average oxygenation values (r^2 = ,613; z = -2,739; p = ,004). In other words, the participants who completed the experimental OOP tasks had low levels of cognitive load, which was represented by the average oxygenation change. Even though this result does not imply causation, it gives an indication for a possible relationship. We observed that the degree, to which the learners automated their OOP skills, became a major factor.

Task automation and schema acquisition, or highly structured knowledge, is very effective in learning a complex skill, like computer programming (Merriënboer and Paas, 1990). Learners are freed from the processing limitations of WM when dealing with previously learned materials. Automation frees up the WM, reduces the cognitive load, and information can be processed automatically without conscious effort (Ericsson and Kinstch, 1995). Familiar and automated tasks can be performed accurately and fluidly. Although previously learned programming tasks can be completed without schema automation, the process is generally slow and awkward (Sweller et al., 1998; Merriënboer and Sweller, 2005).

In this study, learners' unconscious automated skills possibly helped them to solve the experimental problems. This facilitated their cognitive performance by freeing up processing resources that might be devoted to various controlled processes (Merriënboer and Paas, 1990). However, the OOP course and the lecture followed a traditional way for teaching programming to novice learners. The OOP features, along with its paradigm and syntactic details, were initially presented to the learners. Later, a limited number of illustrative examples and solutions demonstrated the use of these features. Finally, the learners were confronted with a number of problems to create new computer programs. This instructional design did not specifically aim to focus on the cognitive dimension. Nonetheless, the learners who completed the experimental tasks unconsciously automated the basic OOP skills to some extent.

They went through the experimental tasks by following the sequence; design, coding, testing and debugging. On the other hand, the learners, who could not complete the tasks, displayed different individual approaches. For example, they directly started coding rather than completing the class hierarchy and inheritance tree. Their bottom-up programming style prevented them from planning and complete understanding of the system. These learners deterred the implementation of "interface D" to the late in the fNIRS sessions (Appendix 1). They did not have enough time to reflect the design requirements to their existing code. As a result, non-automated and ill-organized OOP schema seemed to be an effective factor for noncompleted experimental tasks.

R.Q.-2. What are the subject matter experts' evaluations on the difficulty levels of experimental OOP tasks? Is there any relationship between the cognitive load measurements and the expert level judgments?

The first research question showed that the participants' cognitive load did not change significantly according to their academic achievements, but there was a possible relationship between average oxygenation and the task completion status. However, we were not sure which programming task was more contributing to this result. With this research question, we hoped to see if any oxygenation change existed when executing individual programming tasks. Furthermore, there could be high concentration or disengagement when performing these tasks, which would also cause differences in average oxygenation. Therefore, we explored the instant changes in the oxygenations rather than the average values. Later, these values compared with the subject matter experts' judgments to see whether they showed the same trend. Thus, the experts rated each scale item (column) starting from 1 point to 5 points, which was symbolizing their opinions on difficulty levels of the OOP tasks (Table 2, 5). They were experienced OOP instructors.

Table 5:
OOP expert evaluations on the experimental OOP tasks

Experts	A1	A2	B1	B2	B3	C1	C2	C3	D1	D2	P1	P2
Expert-1	1	2	1	3	3	1	2	3	3	4	3	4
Expert-2	1	2	1	3	3	1	3	3	1	3	2	3
Expert-3	1	1	2	2	3	1	2	3	1	2	3	3
Expert-4	1	1	1	2	2	1	2	2	2	2	3	2
Expert-5	1	2	2	2	2	2	2	2	3	3	2	3
Expert-6	1	1	1	3	4	1	3	4	1	4	3	4
Expert-7	1	2	1	2	3	2	2	3	2	3	3	5
Expert-8	2	2	1	2	4	1	2	4	4	4	4	5
Expert-9	1	3	1	2	2	1	2	2	3	3	3	4
Expert-10	1	1	1	2	2	1	2	2	1	2	3	3
Total	12	19	13	25	31	13	24	31	23	33	32	40
Normalized Value	0,045	0,070	0,049	0,094	0,114	0,049	0,090	0,114	0,086	0,123	0,119	0,147

The experimental protocol included two basic skills. The first was the comprehension of OOP concepts, and the second required their applications to the programming tasks (Table 5). The columns, A1, A2, B1, B2, B3, C1, C2, and C3 were for understanding of the concepts, such as class, inheritance and polymorphism (overloading and overriding).

The D1 and D2 tasks necessitated the implementation of polymorphism in the form of interface. The P1 and P2 tasks directly measured programming skills of the participants. Finally, each rating given to an OOP task was aggregated to a total value in a different row. Thus, it could represent total experts' judgments about a corresponding task. The last row of table included normalized values of the judgments, which were brought to a common scale for the comparisons with average oxygenation changes.

As it is seen in Table 5, the experts regarded B3, C3, D2, P1 and P2 as relatively more difficult than the other conceptual or programming tasks. When observing each participant's session, the experimenter recorded the start and end time of possible events using manual marker of the recording software (Appendix 2). They indicated time intervals of the events and related programming tasks. Additionally, the experimenter recorded the current time of that event, the OOP task in the session, and the detailed explanation of the OOP task. Table 6 presents the oxygenation changes for each participant and experimental task.

Table 6:
Oxygenation changes for each programming task

Participants	A1	A2	B1	B2	B3	C1	C2	C3	D1	D2	P1	P2
P-1					1				1	1		
P-2				1	1						1	
P-3	1		1		1	1						
P-4	1								1		1	
P-5		1			1				1	1		
P-6	1		1		1	1						
P-7		1			1			1				
P-8			1		1				1		1	
P-9				1	1				1	1	1	
P-10				1	1						1	
P-11	1		1	1					1			
Total	4	2	4	4	9	2	0	1	6	3	5	0
Normalized Value	0,100	0,050	0,100	0,100	0,225	0,050	0,000	0,025	0,150	0,075	0,125	0,000

The columns of the Table 6 represent the OOP concepts and programming tasks. Each row belongs to the observations of a participant. The intersection of a row and column with a value of one (1) meant that oxygenation change occurred when the subject was performing the corresponding programming task. The total observations were converted to their normalized values to compare them with the experts' judgments. The normalized values of average oxygenation and expert judgments are plotted in Figure: 2. The analysis of the judgments and oxygenation changes give us that the experts expected the difficulty levels of the experimental tasks be in an increasing fashion. This was from theory to application. The subject matter experts were generally of the opinion that the coding tasks would be more demanding than the conceptual tasks as most of the OOP domain experts would agree. Therefore, P1 and P2 tasks naturally reached the highest peak in the graph (Figure: 2).

However, contrary to the experts' expectations, majority of the oxygenation changes happened during cognitive processing of the conceptual OOP tasks (B3, B2, B1, and D1). Therefore, it would be possible to state that reasoning or problem solving could not occur as the experts foresaw it.

When answering the first research question, we indicated that five participants were not able to complete all of the experimental tasks. They were also unable to meet the programming requirements of D1 and D2 tasks and reflect the changes to their existing code in a specified time.

Regarding the low or high oxygenation as a measure of cognitive load, we have to underline some of the significant factors. The First, as we previously stated, poor task automation and unstructured schema acquisition could possibly result in high cognitive load as well as high oxygenation in WM. The Second, the participants might be overwhelmed by the complexity of conceptual or programming tasks that would cause a decrease in the oxygenation (Izzetoglu et al., 2004; Bunce et al., 2006). However, we saw that they acquired necessary skills for the experimental tasks according to the achievement tests. Since the experimental task were similar to those in the tests, therefore, it is difficult to extrapolate that the participants were affected by the task complexities. Even more interesting, we can attribute the increases in average oxygenation to the high concentration on the tasks, especially for a success (Izzetoglu et al., 2007; León-Carrióna et al., 2010; Ayaz et al., 2012). Because, some of the participants with high achievements had also high average oxygenation levels (Table 2). Therefore, these findings should be interpreted with caution, and further studies are needed to associate the total average oxygenation change with high or low performance in OOP tasks.

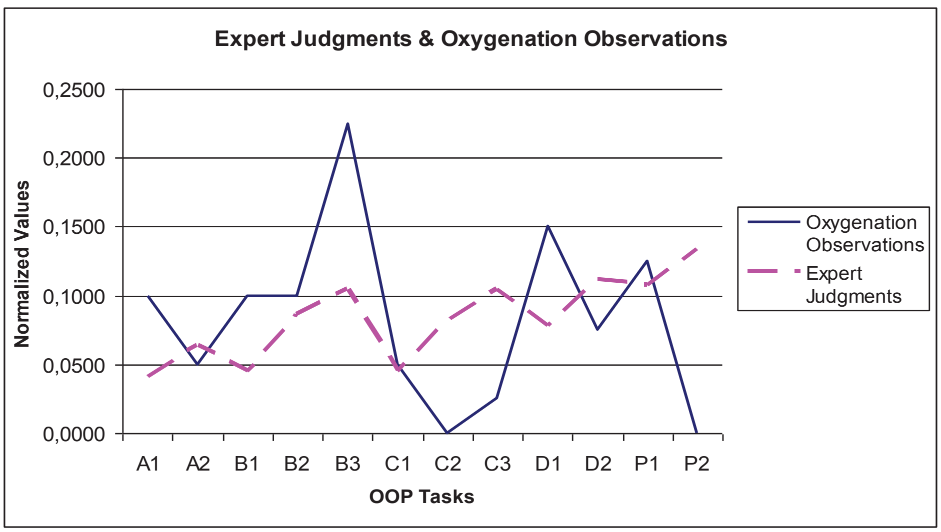

Figure: 2
The comparison of expert judgments and oxygenation changes

These findings can be interpreted as conflicting in its own context. However, we have to take the design of our experiment into consideration as well as our research purpose. It is worth to remind that this experimental study primarily intended to explore the previously learned OOP skills in terms of cognitive load. Since these skills and knowledge were assumed to be acquired, we only wondered if it would be possible to monitor the cognitive performances measured by average oxygenation changes. On the other hand, OOP is composed of highly interactive components requiring various levels of activities. For example, novice learners have to deal with extraneous issues usually stem from development environments, operating systems, etc. The syntax of the language can be an intrinsic issue on its own. Consequently, all of these variables may lead to different types of cognitive load.

Therefore, our research results should be discussed from different point of views. However, this is left to our future studies because of the limitations of this study.

Meanwhile, we would like to emphasize that these findings greatly contributed to the design of our two successive experiments that will be presented in near future. One of them aimed to compare two different instructional Java programming environments. The other one directly focused on the abstraction skills of programmers when executing reverse engineering tasks.

The Human has complicated cognitive processes. How the information is stored, processed both in LTM, and in WM is a determinant factor for an effective cognitive performance. Therefore, instructional design generally determines a meaningful learning and a cognitive performance.

As a result, we regard the findings of this study as the initial experiences of the fNIRS method.

CONCLUSIONS

Teaching OOP is a difficult task, especially to the beginners. First-time learners also find OOP concepts and techniques difficult to grasp. It requires internalization of OOP paradigm and adapting the basic skills of OOP to the problem solving area. Thus, different studies are done on the instructional tools or methods for teaching OOP. Whatever the research topic would be, the cognitive dimension has always been a major factor, and it has played an important role in achieving learning outcomes. Therefore, CLT provided instructional designers with the prescriptions to make information process easier in memory. The strategies covered increasing germane cognitive load as well as reducing intrinsic and extraneous cognitive load. Consequently, how to measure cognitive load of programmers has been one of the issues of instructional designs guided by CLT.

Our study primarily explored how previously learned OOP skills were performed, and how the information was processed in WM. While adopting fNIRS as a relatively new method, we measured the cognitive load of participants when executing programming tasks. The average oxygenation values represented their total cognitive response to a set of OOP tasks. The first research question included two points: (1st) whether the average oxygenation changed according to the academic achievement test, (2nd) if any relationship existed between the task completion status and the average oxygenation change. The second research question investigated the instant changes in the oxygenations. We aimed to find which programming task was more contributing to the oxygenation, and then we compared these results with the experts' judgments.

For the first question, we found that the participants could not reflect their competencies to their cognitive performances in terms of oxygenation change, though they previously learned the experimental OOP tasks. There was not any relationship between the academic achievement tests and the average oxygenation changes.

We believe that the instructional design of the OOP course and supplementary lecture could not provide required cognitive strategies for information processing.

Therefore, this possibly affected the learners' cognitive performance, and their average oxygenation change uncorrelated with the achievement tests. On the other hand, the statistical tests showed a significant and negative correlation between average oxygenation change and task completion status.

The participants, who completed all of the experimental OOP tasks, had also low levels of oxygenations. We observed that they somehow automated their basic OOP skills to a certain extent. This possibly freed up their WM, reduced the cognitive load, and they were able to process information automatically without conscious effort. For the second research question, we found that the subject matter experts expected coding tasks to be more demanding than the conceptual tasks. Should we regard the high average oxygenation as an indicator of high mental effort, we can state that the majority of oxygenation changes occurred during the processing of conceptual tasks. In other words, the problem solving or programming behaviors of the participants could not occur as the experts expected. One of the purposes of the study was to see whether the fNIRS system could have contributions to the research area of instructional technology as a portable and safe technology.

As a result, the fNIRS system was an effective and promising technology for monitoring cognitive tasks both in the classrooms and in the experimental environments. However, our findings should be interpreted with caution regarding the experimental design. We believe that further studies are needed to associate the total average oxygenation change with high or low performances in OOP tasks.

OOP has a high intrinsic load and the instructional designers have to consider this when designing learning environments for programmers. The selected instructional methods and tools should lead to decrease in the extraneous cognitive load while enabling increase in the germane cognitive load.

For example; studies concerning the learners' characteristics together with the fNIR method, i.e. learning styles and personality traits, would bring new perspectives.

The techniques, such as software visualization, animation and concept mapping, can be used to increase germane cognitive load, and their effects can be measured by the fNIRS method. We believe that these suggested topics would bring innovative contributions to the problem area of teaching OOP.

Finally, our paper concludes with an invitation of different studies, which would research different variables within the framework of the fNIRS method.

BIODATA and CONTACT ADRESSESS of AUTHOR

Dr. Murat Paşa UYSAL is Research Asst. Prof. Dr. in TMA. He holds a B.S degree in Electric-Electronic Engineering, M.S in Computer Engineering and Ph.D. in Technology of Education. He teaches computer programming and software related courses. His primary research areas are e-learning and its improvement, instructional methods for computer programming, and computer aided instruction.

Assist. Prof. Dr. Murat Paşa Uysal
Turkish Military Academy, Learning and Research Center, Ankara, TURKEY.
Emails: mpuysal@kho.edu.tr & mpuysal@gmail.com

Acknowledgements: We carried out this research in the laboratories of Modeling and Simulation Research and Development Center, Middle East Technical University. The author would like to thank the members of this center for their support. The author also gratefully acknowledges Dr. Murat Perit Cakir for both his continual support and the data acquisition procedures.

REFERENCES

Artino, A. R. (2008). Cognitive load theory and the role of learner experience: An abbreviated review for educational practitioners. *Association for the Advancement of Computing In Education Journal,* 16, 4, pp. 425-439.

Atkinson, R. C. & Shiffrin, R.M. (1968). *Human memory: A proposed system and its control processes*. In K.W. Spence & J.T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory. Vol. 2, pp. 742-775, New York: Academic Press.

Ayaz, H., Willems, B., Bunce, S., Shewokis, P. A., Izzetoglu, K., Hah, S., Deshmukh, A. & Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. In T. Marek, W. Karwowski & V. Rice (Eds.), *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*, pp. 21-32, CRC Press, Taylor & Francis Group.

Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B. & Onaral, B. (2011). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, doi:10.1016/j.neuroimage.2011.06.023.

Berninger, V. W. & Corina, D. (1998) Making cognitive neuroscience educationally relevant: Creating bidirectional collaborations between educational psychology and cognitive neuroscience. *Educational Psychology Review*, 10, 3, pp. 343–354.

Bozkurt, A., Rosen , A., Rosen, H. & Onaral, B. (2005). A portable near infrared spectroscopy system for bedside monitoring of newborn brain. *BioMedical Engineering OnLine*, doi:10.1186/1475-925X-4-29.

Bunce S.C., Izzetoglu M., Izzetoglu K., Onaral B. & Pourrezai K. (2006). Functional Near-Infrared Spectroscopy: An emerging neuroimaging modality. *IEEE Engineering in Medicine and Biology Magazine*, pp. 54-62.

Campbell, S. R. (2006) Educational Neuroscience: New horizons for research in mathematics education. *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2, pp. 257–264.

Doane, S. M., Sohn, Y. W., McNamara, D. S., & Adams D. (2000). Comprehension-based skill acquisition. *Cognitive Science*, Vol.24, pp. 1-52.

Ericsson, K. A., Kinstch, W. (1995). Long-term working memory. *Psychological Review*, Vol. 102, No. 2, pp. 211-245.

Holtzer, R., Mahoney, J.R., Izzetoglu, M., Izzetoglu, K., Onaral B. & Verghese J. (2011). fNIRS study of walking and walking while talking in young and old individuals. Journal *of Gerontology*, doi:10.1093/gerona/glr068.

Izzetoglu, K., Bunce S., Onaral B. & Chance B. (2004). Functional brain imaging using near-infrared during cognitive tasks. *International Journal of Human-Computer Interaction*, 17 (2), 211-227.

Izzetoglu, M., Izzetoglu, K., Bunce S., Ayaz H., Devaraj A. & Onaral B. (2005). Functional near-infrared neuroimaging. *IEEE Transactions Neural Systems and Rehabilitation Engineering*, 13, pp. 153-159.

Izzetoglu M., Bunce S. C., Izzetoglu K., Onaral B. & Pourrezai K. (2007). Functional brain imaging using near-infrared technology: Assessing cognitive activity in real-life situations. *IEEE Engineering in Medicine and Biology Magazine*, pp. 36-44.

Jobsis F.F. (1977). Noninvasive infrared monitoring of cerebral and myocardial sufficiency and circulatory parameters. *Science*, 198, pp. 1264 -1267.

Griffin, S. & Case, R. (1997) Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, 3, 1, pp. 1-49.

Kirschner, P. A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, pp. 75-86.

Lenf R.D., Orihuella-Espina F., Elwell C.E., Athanasiou T., Delpy D.T., Darzi A. & Yang G.Z. (2011). Assessment of the cerebral cortex during motor task behaviours in adults: A systematic review of functional near infrared spectroscopy (fNIRS) studies. *Neuroimage*, 54, pp. 2922-2936.

León-Carrióna J., Izzetoglu M., Izzetoglu K., Martín-Rodrígueza J.F., Damas-López J., Barroso J.M.M. & Morales M. R. D. (2010). Efficient learning produces spontaneous neural repetition suppression in prefrontal cortex. *Behavioural Brain Research*, 208, pp. 502-508.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, pp. 43-52.

Merriënboer, J.G., & Paas, G.W.C. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior*, Vol.6, pp. 273-289.

Modsimmer, (2010). The seminar on Functional Near-Infrared (fNIR) Spectroscopy system. Retrieved from "http://www.modsim.metu.edu.tr/index.php?option=com_content&view=article&id=360&Itemid=50" on November 12, 2010.

OECD (2002). Understanding the brain: Towards a new learning science (Paris, OECD).

Schneider, W. & Detweiler, M. (1987). A connectionist / control architecture for working memory. The *Psychology of Learning and Motivation*, Vol.21, pp. 53-119.

Shiffrin, R.M. (1976). Capacity limitations in information processing, attention and memory. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes*. Vol. 4, pp. 177-236, Hillsdale, N.J.: Erlbaum.

Sohn,Y.W., & Doane, S.M. (1997). Cognitive constraints on computer problem-solving skills. *Journal of Experimental Psychology*: *Applied*, Vol.3, pp. 288-312.

Sweller, J., Van Merriënboer, Jeroen, J.G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, Vol.10, pp. 251-295.

Szucs, D., Soltész, F., Jármi, E & Csépe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: An electro-encephalography study*. Behavioral and Brain Functions*, 3, 23, doi:10.1186/1744-9081-3-23.

Xinogalos, S. (2010). Guidelines for designing and teaching an effective object-oriented design and programming course. *Advance Learning*, pp. 397-422.

Van Nes, F. & De Lange, J. (2007). Mathematics education and neurosciences: Relating spatial structures to the development of spatial sense and number sense. *The Montana Mathematics Enthusiast*, 4:2, pp. 210–229.

Van Nes, F. (2011). Mathematics education and neurosciences: Towards interdisciplinary insights into the development of young children's mathematical abilities. *Educational Philosophy and Theory*, Vol. 43, No. 1, 2011. doi: 10.1111/j.1469-5812.2010.00710.x

Villringer A. & Chance B. (1997) Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, Vol.20, pp. 435-42.

Waugh, N.C., & Norman, D.A. (1968). The measure of interference in primary memory. *Journal of Verbal Learning and Verbal Behavior*, Vol. 7, pp. 617-626.

White, G. & Sivitanides M. (2005). Cognitive differences between procedural programming and object oriented programming. *Journal of Information Technology and Management*, Vol.6, Issue 4, pp. 330-350.

Y. W., Soan & S.M., Doane (2003) Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory and Cognition*, 31, 3, pp.458-466.

Appendix-1:
The Experimental Protocol and OOP expert evaluation scale

A research study on "novice object-oriented programmers" will be carried out. The difficulty levels of the concepts and programming tasks need to be determined by subject-matter experts. We would like to thank you for your contributions to this study.

1–Very easy; 2–Easy; 3–Moderate; 4–Difficult 5–Very difficult;

Relationships
A is a base class; B and C is a subclass of A; D is an interface implemented by B and C

	Difficulty Levels	1	2	3	4	5
Class A						
A1)	m_a1 and m_a2 are integer type private attributes					
A2)	DisplayName( ) is a method of class A, and it returns the "A" as a string value					
Class B						
B1)	m_b1 and m_b2 are integer and double type private attributes					
B2)	Class B overrides displayName( ) method of A and it returns the "B" as a string value					
B3)	Class B has a overloaded public method named as multiply( ). The first overloaded one returns an integer value and it has a signature as (int, int). The second overloaded method returns a double value and it has a signature as (double, double). They both make multiplication operation					
Class C						
C1)	m_c1 and m_c2 are private integer and double type attributes					
C2)	Class C overrides displayName( ) method of A and it returns "C" as a string value					
C3)	Class C has an overloaded public method named as sum ( ). The first overloaded method returns an integer value and it has a signature as (int, int). The second overloaded method returns a double value and it has a signature as (double, double). They both make summation operation					
Interface D						
D1)	D has a method public void method named as reset ( ), which has no parameters.					
D2)	Class B and C implements the interface D, and the implemented reset() method assigns the "0" value to the class' own attributes					
Programming Tasks						
P1)	Write a Java application named as "Test.java". Instantiate the B and C classes. Then B and C classes should perform the sum and multiplication arithmetic operations respectively. You should be able to observe the outputs of your program.					
P2)	Change the B and C class definitions so that they implement the interface D.					

Appendix-2:
COBI data collection software suit